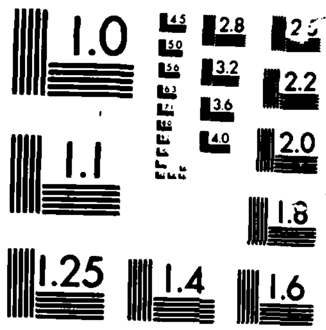
AD-A166 179 STATE ESTIMATION FOR COX PROCESSES WITH UNKNOWN LAW: 1/1
PARAMETRIC MODELS(U) JOHNS HOPKINS UNIV BALTIMORE MD
DEPT OF MATHEMATICAL SCIENCES A F KARR NOV 85
UNCLASSIFIED AFOSR-TR-86-0020 F/G 12/1 NL



MICROCOPY RESOLUTION TEST CHART

AD-A166 179

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

(2)

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION: Unclassified

AD-A166179

1b. RESTRICTIVE MARKINGS:

2a. SECURITY CLASSIFICATION AUTHORITY:

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE:

3. DISTRIBUTION/AVAILABILITY OF REPORT: Approved for public release; distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S):

5. MONITORING ORGANIZATION REPORT NUMBER(S): AFOSR-TR- 86-0020

6a. NAME OF PERFORMING ORGANIZATION: Johns Hopkins University

6b. OFFICE SYMBOL *(If applicable)*:

7a. NAME OF MONITORING ORGANIZATION: AFOSR/NM

6c. ADDRESS *(City, State and ZIP Code)*: Baltimore, MD 21218

7b. ADDRESS *(City, State and ZIP Code)*: Bolling AFB, D.C. 20332-6448

8a. NAME OF FUNDING/SPONSORING ORGANIZATION: AFOSR

8b. OFFICE SYMBOL *(If applicable)*: NM

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER: ~~F49620-82-C-0029~~ AFOSR-82-0029

8c. ADDRESS *(City, State and ZIP Code)*: Bolling AFB, D.C. 20332-6448

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.
61102F	2304	A5	

11. TITLE *(Include Security Classification)*: State Estimation for Cox Processes with Unknown Law: Parametric Models

12. PERSONAL AUTHOR(S): Alan F. Karr

13a. TYPE OF REPORT: Technical

13b. TIME COVERED: FROM ________ TO ________

14. DATE OF REPORT *(Yr., Mo., Day)*: November 1985

15. PAGE COUNT: 16

16. SUPPLEMENTARY NOTATION:

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*: Cox process, point process, state estimation for point processes, parametric statistical model, maximum likelihood estimator, EM algorithm

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*

Let (N_i, M_i) be i.i.d. copies of a Cox pair (N,M), with the Cox processes N_i, but not the directing measures M_i, observable. Suppose that the probability law of (N,M) belongs to a finite-dimensional parametric family (P_θ) but is otherwise unknown. Approximations are derived for state estimators $E_\theta[M_{n+1}|F^{N_{n+1}}]$, in which the unknown value θ is replaced by the maximum likelihood estimator $\hat{\theta}$ based on $N_1,\ldots,N_n$, leading to pseudo-state estimators $E_{\hat{\theta}}[M_{n+1}|F^{N_{n+1}}]$. Under standard smoothness and regularity assumptions $n^{1/2}$ times the difference between the true and pseudo-state estimators converges in distribution to a Gaussian random measure, with respect to the variation norm topology. Computation of maximum likelihood estimators by the EM algorithm is discussed in general and for two specific examples.

DTIC FILE COPY

DTIC ELECTE APR 0 1 1986 S D

D

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT: UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION: Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL: Major Brian W. Woodruff

22b. TELEPHONE NUMBER *(Include Area Code)*: (202) 767-5027

22c. OFFICE SYMBOL: NM

DD FORM 1473, 83 APR — EDITION OF 1 JAN 73 IS OBSOLETE.

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

DISCLAIMER NOTICE

AFOSR-TR- 86-0020

DEPARTMENT OF MATHEMATICAL SCIENCES
The Johns Hopkins University
Baltimore, Maryland 21218

State Estimation for Cox Processes with Unknown Law: Parametric Models*

Alan F. Karr

Technical Report No. 449
November, 1985

*Research supported by the Air Force Office of Scientific Research, AFSC, USAF, under grant number ~~82-0029C~~. The United States Government is authorized to reproduce and distribute reprints for governmental purposes.

AFOSR-82-0029

Accesion For		
NTIS CRA&I	☑	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1	23	

ABSTRACT

Let (N_i, M_i) be i.i.d. copies of a Cox pair (N,M), with the Cox processes N_i, but not the directing measures M_i, observable. Suppose that the probability law of (N,M) belongs to a finite-dimensional parametric family (P_θ) but is otherwise unknown. Approximations are derived for state estimators $E_\theta[M_{n+1}|F^{N_{n+1}}]$, in which the unknown value θ is replaced by the maximum likelihood estimator $\hat{\theta}$ based on $N_1,\ldots,N_n$, leading to pseudo-state estimators $E_{\hat{\theta}}[M_{n+1}|F^{N_{n+1}}]$. Under standard smoothness and regularity assumptions $n^{\frac{1}{2}}$ times the difference between the true and pseudo-state estimators converges in distribution to a Gaussian random measure, with respect to the variation norm topology. Computation of maximum likelihood estimators by the EM algorithm is discussed in general and for two specific examples.

Keywords:

AMS 1980 subject classifications: primary 60G55, 60G57; secondary 62F10, 62M20, 93E10

Key words and phrases: Cox process, point process, state estimation for point processes, parametric statistical model; maximum likelihood estimator; EM algorithm

037

1. Introduction

A point process $N = \Sigma \varepsilon_{X_i}$ and finite random measure M on a locally compact Hausdorff space E form a Cox pair if they are defined over the same probability space and if conditional on $F^M = \sigma(M(A) : A \in \mathcal{E})$ (here $\mathcal{E}$ is the Borel σ-algebra on E), N is a Poisson process with mean measure M. We say also that N is a Cox process directed by M and that M is the directing measure of N. For expository treatment see [4, 8]; [10] emphasizes engineering applications with $E = \mathbb{R}_+$.

In many applications the directing measure M represents an underlying random mechanism of greater physical importance than N, but only the latter is observable. For example, in optical communication [10], the observed process of photon counts is (modeled as) a Cox process on $\mathbb{R}_+$ with directing measure $M(dt) = X_t dt$, where $X \geq 0$ is a signal process. State estimation for the directing measure M of a Cox process N, given observation of N over a subset A of E, is the problem of optimal reconstruction, realization-by-realization, of the unobserved measure $M(\omega)$. Assuming that $E[M(E)] < \infty$, optimal state estimators are conditional expectations $E[M(f)|F^N(A)]$, where $M(f) = \int f dM$ is the integral of the (nonnegative, continuous, compactly supported) function f with respect to M and $F^N(A) = \sigma(N(B) : B \subseteq A)$ is the σ-algebra corresponding to (complete, uncorrupted) observation of N over A (when $A = E$ we write simply F^N).

Given that the probability law of M (or that of N -- they determine each other; cf. [8]) is known, the state estimation problem is solved in essentially full generality. The most general form of the solution [7] is

$$E[M(f)|F^N(A)] = E[e^{-M^{(\mu)}(A)}M^{(\mu)}(f)]/E[e^{-M^{(\mu)}(A)}]\Big|_{\mu=N_A}, \tag{1.1}$$

where the $M^{(\mu)}$ are Palm processes of M and $N_A(\cdot) = N(\cdot \cap A)$ is the observation of N. Definition of Palm processes is intricate; heuristically, for $\mu = \sum_{i=1}^{r} \varepsilon_{x_i}$,

$$P(M^{(\mu)} \in \Gamma) = E[1(M \in \Gamma) \prod_{i=1}^{r} M(dx_i)]/E[\prod_{i=1}^{r} M(dx_i)]$$

in the sense of Radon-Nikodym derivatives. See [4, 7, 8] for more detailed explanation.

In this paper we work at a level of reduced generality. Specifically, when

$$M(dx) = X(x)\nu(dx) \tag{1.2}$$

with X a nonnegative, measurable process and ν a fixed, finite measure on E, (1.1) becomes [5]

$$E[M(f)|F^N(A)] = \frac{\int L(d\eta)\exp[-\eta(A) + \int_A \log(d\eta/d\nu)dN]\eta(f)}{\int L(d\eta)\exp[-\eta(A) + \int_A \log(d\eta/d\nu)dN]}, \tag{1.3}$$

where L is the probability law of M. In the still more special case that $M(dx) = Y\nu(dx)$, where Y is a positive random variable (with distribution function F), N is called a mixed Poisson process and (1.3) becomes

$$E[M(f)|F^N(A)] = \frac{\int F(dt)e^{-t\nu(A)}t^{N(A)+1}}{\int F(dt)e^{-t\nu(A)}t^{N(A)}}\nu(f). \tag{1.4}$$

This paper continues our study [6, 7, 8] of the problem of combined statistical inference and state estimation, i.e., state estimation when the probability law is unknown. Our formulation is as follows. Let (N_1, M_1), $(N_2, M_2), \ldots$ be i.i.d. copies of a Cox pair (N, M) with unknown law. Suppose that only the N_i are observable, and are observed, one after another, over

(for simplicity) the entire space E. To goal is to construct and analyze approximations to the "true" (but uncomputable, because to do so entails knowledge of the probability P) state estimators $E[M_{n+1}|F^{N}_{n+1}]$. Our general approach invokes the <u>principle of separation</u> used for some time in electrical engineering. By (1.1), $E[M_{n+1}|F^{N}_{n+1}] = H(L, N_{n+1})$, where L is the law of the N_i and M, defined by the right-hand side of (1.1), is a functional of known -- albeit complicated -- form (shown more clearly by (1.3)). Given an estimator $\hat{L}$ of L based on $N_1, \ldots, N_{n+1}$, we form a <u>pseudo-state estimator</u> by substituting $\hat{L}$ for L: $\hat{E}[M_{n+1}|F^{N}_{n+1}] = H(\hat{L}, N_{n+1})$. The main results here, as in [6, 7], pertain to asymptotic behavior of differences $\hat{E}[M_{n+1}|F^{N}_{n+1}] - E[M_{n+1}|F^{N}_{n+1}]$.

In [6] the N_i were mixed Poisson processes; by exploiting special structure intensively we developed pseudo-state estimators based on nonparametric estimators of ν and F (cf. (1.4)), and were able to show that the error processes $\{n^{\frac{1}{2}}(\hat{E}[M_{n+1}|F^{N}_{n+1}] - E[M_{n+1}|F^{N}_{n+1}])\}$, regarded as signed random measures on E, converge to a mixture of certain Gaussian processes. This result is essentially optimal in regard to rate of convergence: pseudo-state estimators cannot approximate true at a rate faster than L itself (or, in this case, objects defining L) can be estimated. A streamlined presentation of the mixed Poisson case is given in [8, Section 7.4].

By contrast, the unrestricted, nonparametric setting of [7] engendered rather severe consequences. Pseudo-state estimators were constructed by estimating the <u>Palm distributions</u> of M (the laws of the Palm processes $M^{(\mu)}$ in (1.1)), which stand related as Cox pairs to those of N; the latter are estimated from observation of the N_i. Asymptotic normality of these estimators has been demonstrated only in an integrated form [7, Theorem 4.2], making it possible to show only that for $\delta > 0$,

$$n^{\frac{1}{4}-\delta}(\hat{E}[M_{n+1}|F^{N}_{n+1}] - E[M_{n+1}|F^{N}_{n+1}]) \to 0$$

in the sense of an appropriate (for fixed but suppressed functions f) form of L^2-convergence. This rate of $n^{-(\frac{1}{4})+\delta}$ is <u>much</u> worse than the rate $n^{-(\frac{1}{2})+\delta}$ for the mixed Poisson case.

Here we take another tack, analysing the problem in a <u>parametric</u> setting but imposing no structural restrictions beyond (1.2) on the directing measure. Thus, let Θ be a relatively compact, open subset of $\mathbb{R}$ (extension to the multidimensional case is straightforward). Suppose that for each $\theta \in \Theta$ we have a probability measure P_θ under which $(N_1, M_1), (N_2, M_2), \ldots$ are i.i.d. copies of a Cox pair (N, M), and that the directing measure M has P_θ-law $L_\theta^M(\cdot)$. The assumptions introduced here are presumed in force throughout the paper.

ASSUMPTIONS 1.1. a) The statistical model $\{P_\theta : \theta \in \Theta\}$ is dominated: there is a probability measure Q on the set $\underline{M}$ of finite measure on E with respect to which there exist density functions f_θ satisfying

$$L_\theta^M(d\nu) = f_\theta(\nu)Q(d\nu). \tag{1.5}$$

b) $Q(\underline{M}_d) = 1$, where $\underline{M}_d$ is the set of diffuse measures (i.e., those without atoms) in $\underline{M}$.

c) There is $\nu^* \in \underline{M}_d$ such that $\nu << \nu^*$ for Q-almost all ν.

d) $\{f_\theta(\cdot) : \theta \in \Theta\}$ is identifiable: $\theta \neq \hat{\theta}$ implies that $Q(\{\nu : f_\theta(\nu) \neq f_{\hat{\theta}}(\nu)\}) >$

The first and fourth are quite standard in parametric statistical inference; see for example [1]. The second and third are more specific, ensuring that the Cox processes N_i be simple (without multiple points) and placing us in the context of (1.2) - (1.3), respectively.

Let

$$K(\nu,\mu) = \exp[\nu^*(E)-\nu(E) + \int_E \log(d\nu/d\nu^*)d\mu] \tag{1.6}$$

be (as a function of $\mu \in \underline{M}_s$, the set of simple point measures on E) the Radon-Nikodym derivative of the law of the Poisson process with mean measure $\nu \ll \nu^*$ with respect to that of the Poisson process with mean measure ν^*; cf. [8, Section 6.2] for details. In view of Assumption 1.1a) the N_i have P_θ-density function

$$g_\theta(\mu) = \int Q(d\nu) f_\theta(\nu) K(\nu,\mu) \tag{1.7}$$

with respect to the ν^*-Poisson law. Since the laws of a Cox process and its directing measure determine each other, Assumption 1.1d) implies that $\{g_\theta(\cdot)\}$ is identifiable.

Corresponding to observation of $N_1,\dots,N_n$ is the n-sample log-likelihood function

$$L_n(\theta) = \sum_{i=1}^{n} \log \int Q(d\nu) f_\theta(\nu) K(\nu,N_i), \tag{1.8}$$

which plays a major role in our analysis. Note that (1.8) exhibits our model as a mixture model (cf. [9] for estimation procedures for other mixture models); however we make little overt use of this property.

The remainder of the paper is organized in the *following* manner. Section 2 deals with existence and computation of maximum likelihood estimators $\hat{\theta} = \hat{\theta}_n$. Because only the Cox processes N_i are observable we must deal with the cumbersome log-likelihood functions L_n of (1.8) rather than the potentially much more tractable log-likelihood functions

$$\Lambda_n(\theta) = \sum_{i=1}^{n} \log f_\theta(M_i) \tag{1.9}$$

associated with the M_i; the latter -- because the directing measures are unobservable -- cannot be calculated from the observations. The EM *algorithm* of [2] is a key computational device. In Section 3 we present asymptotic properties of differences

$$E_{\hat{\theta}_n}[M_{n+1}|F^{N_{n+1}}] - E_\theta[M_{n+1}|F^{N_{n+1}}] \tag{1.10}$$

between pseudo- and true state estimators; we impose rather standard regularity conditions on $\{f_\theta\}$ and do obtain optimal rates of convergence. Finally, Section 4 illustrates our results with two specific examples.

To conclude this section we note explicitly that all our conditions are imposed on the M-density functions f_θ rather than the N-density functions g_θ; this is because in theory as well as applications the former constitute the model and the latter are simply derived via (1.7).

2. Existence and computation of maximum likelihood estimators

Our initial concern is existence of maximizers of the log-likelihood functions L_n of (1.8), then whether these estimators fulfill suitable likelihood equations, and finally how solutions of likelihood equations can be calculated effectively. We impose only conditions slightly stronger (typically by addition of uniformity) than needed to derive analogous properties of the uncomputable M-log-likelihood functions Λ_n of (1.9).

We begin with existence; whenever no confusion seems to arise we suppress dependence of estimators on the sample size n.

PROPOSITION 2.1. Assume that the family of functions $\theta \to f_\theta(\nu), \nu \in \underline{M}_d$, is equicontinuous on $\bar{\Theta}$ (the closure of Θ). Then for each n there exists a

maximum likelihood estimator $\hat{\theta}$ for the likelihood function $L_n(\theta)$ of (1.8).

PROOF. It suffices to observe that L_n is continuous on $\bar{\Theta}$, which holds if the density functions $g_\theta(\mu)$ of the N_i, defined by (1.7), are continuous in θ for fixed μ, which is nearly immediate. Given μ and $\epsilon > 0$, let $\delta > 0$ be such that $|\theta - \bar{\theta}| < \delta$ implies

$$\sup\{|f_\theta(\nu) - f_{\bar{\theta}}(\nu)| : \nu \in \underline{M}_d\} < \epsilon / \int Q(d\nu) K(\nu,\mu).$$

Then whenever $|\theta - \bar{\theta}| < \delta$,

$$|g_\theta(\mu) - g_{\bar{\theta}}(\mu)| \le \int Q(d\nu)\, |f_\theta(\nu) - f_{\bar{\theta}}(\nu)|\, K(\nu,\mu) \le \epsilon. \quad \square$$

Under further conditions the maximum likelihood estimator $\hat{\theta}$ fulfills the likelihood equation

$$0 = L_n'(\theta) = \sum_{i=1}^{n} \frac{\int Q(d\nu) f_\theta'(\nu) K(\nu, N_i)}{\int Q(d\nu) f_\theta(\nu) K(\nu, N_i)}, \tag{2.1}$$

the second equality here will be confirmed momentarily. In (2.1) and below, all derivatives are with respect to θ.

PROPOSITION 2.2. Assume that $\theta \to f_\theta(\nu)$ is continuously differentiable on Θ for each ν and that

$$\sup\{|f_\theta'(\nu)| : \theta \in \Theta, \nu \in \underline{M}_d\} < \infty. \tag{2.2}$$

Then for each n the maximum likelihood estimator satisfies

$$L_n'(\hat{\theta}) = 0. \tag{2.3}$$

PROOF. We need to establish differentiability of $L_n(\cdot)$, which follows from

that of $g_\theta(\mu)$ for fixed μ. By virtue of (2.2), Fubini's theorem implies that (for $h > 0$, say)

$$\begin{aligned} h^{-1}[g_{\theta+h}(\mu) - g_\theta(\mu)] &= \int Q(d\nu) \left[\int_\theta^{\theta+h} f_\xi'(\nu)\, d\xi\right] K(\nu,\mu) \\ &= h^{-1} \int_\theta^{\theta+h} \left[\int Q(d\nu) f_\xi'(\nu) K(\nu,\mu)\right] d\xi \\ &\to \int Q(d\nu) f_\theta'(\nu) K(\nu,\mu) \end{aligned}$$

as $h \to 0$, where the last step is by the dominated convergence theorem, which is applicable by (2.2). $\square$

REMARKS. 1) Observe that (2.2) is stronger than the equicontinuity postulated in Proposition 2.1; indeed, assuming uniformly bounded derivatives is arguably the most common way of engendering equicontinuity.

2) In the proof of Proposition 2.1 we have justified interchanging integrals and derivatives in the definition of $g_\theta(\mu)$ and hence within $L_n(\theta)$ as well. Thus the second equality in (2.1) follows.

Solution of the nonlinear equation (2.1) (a system of equations in the multidimensional case) is -- as the examples in Section 4 illustrate -- cumbersome at best or even impossible. The EM algorithm of [2], fortunately, is an effective method for solving this equation iteratively; it was designed for partial observation contexts akin to ours. Its applicability is established in the following result. For each n, let $G_n = \bigvee_{i=1}^{n} F^{N_i}$ and recall from (1.9) that $\Lambda_n(\theta)$ is the unobservable log-likelihood function for $N_1, \ldots, N_n$.

PROPOSITION 2.3. Suppose that the hypotheses of Proposition 2.2 are fulfilled. Then for $\bar{\theta}$, $\theta^* \in \Theta$,

$$\sum_{i=1}^{n} \frac{\int Q(dv) f_{\tilde\theta}(v) K(v,N_i) [f'_{\theta^*}(v)/f_{\theta^*}(v)]}{\int Q(dv) f_{\tilde\theta}(v) K(v,N_i)} \tag{2.4}$$

$$= \frac{d}{d\theta} E_{\tilde\theta}[\Lambda_n(\theta)|G_n]\Big|_{\theta=\theta^*}$$

PROOF. The derivation of (1.3) (see [5, 8]) establishes that, more generally,

$$E_\theta[G(N)|F^N] = \frac{\int Q(dv) f_\theta(v) \exp[-v(E)+\int_E \log(dv/dv^*)dN] G(v)}{\int Q(dv) f_\theta(v) \exp[-v(E)+\int_E \log(dv/dv^*)dN]}$$

$$= \frac{\int Q(dv) f_\theta(v) K(v,N) G(v)}{\int Q(dv) f_\theta(v) K(v,N)} .$$

By independence of the pairs (N_i, M_i),

$$E_{\tilde\theta}[\Lambda_n(\theta)|G_n] = \sum_{i=1}^{n} E_{\tilde\theta}[\log f_\theta(M_i)|F^{N_i}] \tag{2.5}$$

$$= \sum_{i=1}^{n} \frac{\int Q(dv) f_{\tilde\theta}(v) K(v,N_i) \log f_\theta(v)}{\int Q(dv) f_{\tilde\theta}(v) K(v,N_i)} .$$

In view of Proposition 2.2 this last expression is differentiable in θ; evaluating the derivative at θ^* produces (2.4). □

The EM algorithm then proceeds in the following manner. (The sample size is fixed.)

1) Beginning with an initial or previous estimator $\hat\theta^0$, form the <u>pseudo-log -likelihood function</u> $\theta \to E_{\hat\theta^0}[\Lambda_n(\theta)|G_n]$. Except that it pertains to a log-likelihood function rather than a directing measure, this is the same kind of approximate reconstruction of an unobservable random variable forming the main subject of the paper. At the moment, though, our interest is

focussed on statistical inference rather than state estimation.

2) Calculate a new estimator $\hat\theta^1$ maximizing the pseudo-log-likelihood function, by solving the (pseudo-) likelihood equation $(d/d\theta)E_{\hat\theta^0}[\Lambda_n(\theta)|G_n] = 0$. By (2.4) this equation is equivalent to

$$\sum_{i=1}^{n} \frac{\int Q(dv) f_{\hat\theta^0}(v) K(v,N_i) [f'_\theta(v)/f_\theta(v)]}{\int Q(dv) f_{\hat\theta^0}(v) K(v,N_i)} = 0. \tag{2.6}$$

3) Repeat the iteration until the sequence of estimators $\hat\theta^k$ converges. (We prove below that it does, under mild restrictions; in practice, of course, "convergence" means that $|\hat\theta^{k+1} - \hat\theta^k|$ is within some tolerance.) The limit evidently satisfies (2.1) and hence is the n-sample maximum likelihood estimator.

We now present the relevant theory.

THEOREM 2.4. Let the sample size n be fixed; assume that the hypotheses of Proposition 2.2 hold.

a) Let $\tilde\theta$ be fixed and let θ^* be a maximizer of the pseudo-log-likelihood function $\theta \to E_{\tilde\theta}[\Lambda_n(\theta)|G_n]$. Then

$$L_n(\theta^*) \geq L_n(\tilde\theta), \tag{2.7}$$

with the inequality strict unless $\tilde\theta$ is a maximizer of L_n.

b) Suppose that $\{f_\theta(v)\}$ is an exponential family. Then for every initializing value $\hat\theta^0 \in \Theta$ the resultant iterate sequence $(\hat\theta^k)$ generated by the EM algorithm converges to the maximum likelihood estimator $\hat\theta$.

That is, unless the current value is a maximizer of the n-sample likelihood function L_n, the EM algorithm will yield an increase in the likelihood

function. Under b), the maximum likelihood estimator $\hat{\theta}$ is unique, and convergence to it is guaranteed regardless of the initializing value. We sketch the proof of selected aspects of the theorem that are simple and instructive in our case; the main points, especially concerning convergence are known -- see [11, 12].

PROOF OF THEOREM 2.4. a) From (2.5) and the definition of θ^* as maximizer of $E_{\tilde{\theta}}[\Lambda_n(\cdot)|G_n]$,

$$\begin{aligned}
\sum_{i=1}^{n} \frac{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i) \log f_{\tilde{\theta}}(\nu)}{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i)}
&= E_{\tilde{\theta}}[\Lambda_n(\tilde{\theta})|G_n] \\
&\le E_{\tilde{\theta}}[\Lambda_n(\theta^*)|G_n] \\
&= \sum_{i=1}^{n} \frac{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i) \log f_{\theta^*}(\nu)}{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i)} \\
&= \sum_{i=1}^{n} \left\{ \frac{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i) \log[f_{\theta^*}(\nu)/f_{\tilde{\theta}}(\nu)]}{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i)} + \frac{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i) \log f_{\tilde{\theta}}(\nu)}{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i)} \right\} \\
&\le \sum_{i=1}^{n} \log \left\{ \frac{\int Q(d\nu) f_{\theta^*}(\nu) K(\nu,N_i)}{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i)} \right\} + \sum_{i=1}^{n} \frac{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i) \log f_{\tilde{\theta}}(\nu)}{\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i)}
\end{aligned}$$

by Jensen's inequality. Hence

$$\begin{aligned}
0 &\le \sum_{i=1}^{n} \log\left(\int Q(d\nu) f_{\theta^*}(\nu) K(\nu,N_i)\right) - \sum_{i=1}^{n} \log\left(\int Q(d\nu) f_{\tilde{\theta}}(\nu) K(\nu,N_i)\right) \\
&= L_n(\theta^*) - L_n(\tilde{\theta}).
\end{aligned}$$

Equality can hold only if Jensen's inequality becomes equality as well, which by identifiability (Assumption 1.1d)) implies that $\theta^* = \tilde{\theta}$.

b) Compactness of $\bar{\Theta}$ and a) imply that every subsequence of $(\hat{\theta}^k)$ contains a further subsequence converging to a maximizing value of the likelihood function L_n. Because $\{f_\theta(\nu)\}$ is an exponential family (see [1,3] for expository discussion) the same is true of $\{f_\theta(\nu)K(\nu,\mu)\}$ and hence [12, Theorem 3 and subsequent discussion] applies to complete the proof of b). ∎

Whether the EM algorithm is effective (it may be that (2.6) is no easier to solve than (2.1)) depends on $\{f_\theta\}$ and has no universal answer. The examples in Section 4 are cases in which it is effective.

3. Asymptotic properties of pseudo-state estimators

We continue in the setting of *i.i.d. Cox pairs* (N_i, M_i), with the Cox processes observable but the directing measures not. To reconstruct M_{n+1} from observation of N_{n+1}, if the true parameter value θ were known we would calculate the state estimator

$$E_\theta[M_{n+1}|F^{N_{n+1}}] = \frac{\int Q(d\nu) f_\theta(\nu) K(\nu,N_{n+1}) \nu(\cdot)}{\int Q(d\nu) f_\theta(\nu) K(\nu,N_{n+1})} . \qquad (3.1)$$

When θ is unknown, we must -- as discussed in Section 1, resort to the pseudo-state estimator

$$E_{\hat\theta}[M_{n+1}|F^{N_{n+1}}] = \frac{\int Q(dv)\ell_{\hat\theta}(v)K(v,N_{n+1})v(\cdot)}{\int Q(dv)\ell_{\hat\theta}(v)K(v,N_{n+1})}, \tag{3.2}$$

where $\hat\theta$ is the n-sample maximum likelihood estimator of θ. Note a subtle but very convenient change from the description in Section 1: even though good statistical practice dictates otherwise, we have not used the additional data N_{n+1} in forming the estimator $\hat\theta$. The useful consequence is that the two random terms in (3.2), namely $\hat\theta$ and N_{n+1}, are independant, rather than asymptotically independent, as they would be if $\hat\theta$ were the $(n+1)$-sample maximum likelihood estimator. Observe also that (3.1) and (3.2) exhibit $E_\theta[M_{n+1}|F^{N_{n+1}}]$ and $E_{\hat\theta}[M_{n+1}|F^{N_{n+1}}]$ as random measures in the sense of [4].

We now present our main results, which concern asymptotic behavior of the difference between the true and pseudo-state estimators; we begin with a weak consistency theorem. In order to simplify notation within conditional expectations we write N_{n+1} for $F^{N_{n+1}}$; keep in mind as well that dependence of $\hat\theta$ on n is *suppressed*.

THEOREM 3.1. Assume that the hypotheses of Proposition 2.1 are satisfied, that $E_\theta[M(E)] < \infty$ and that $\int Q(dv)v(E) < \infty$. Then as $n \to \infty$,

$$\| E_{\hat\theta}[M_{n+1}|N_{n+1}] - E_\theta[M_{n+1}|N_{n+1}] \| \to 0 \tag{3.3}$$

in the sense of convergence in P_θ-probability, where $\|\cdot\|$ denotes the total variation norm on the space of finite measures on E.

PROOF. Let $\tilde Q$ be the probability law of the Poisson process with E with mean measure v^*; with respect to $\tilde Q$ the N_i have P_θ-density g_θ given by (1.7).

We show first that the maximum likelihood estimators $\hat\theta$ are <u>strongly</u>

consistent: $\hat\theta \to \theta$ P_θ-almost surely as $n \to \infty$. For this we appeal to [3, Theorem I.4.3], and must verify that

a) $\theta \to g_\theta(\mu)$ is continuous for each μ;

b) for each $\theta \in \Theta$ and $\gamma > 0$,

$$\inf_{|\tilde\theta-\theta|>\gamma} \int [g_\theta(\mu)^{1/2} - g_{\tilde\theta}(\mu)^{1/2}]^2 \tilde Q(d\mu) > 0; \tag{3.4}$$

c) for each $\theta \in \Theta$,

$$\lim_{\delta\to0} \int \sup_{|h|\le\delta} [g_{\theta+h}(\mu)^{1/2} - g_\theta(\mu)^{1/2}]^2 \tilde Q(d\mu) = 0. \tag{3.5}$$

The first is an immediate consequence of the equicontinuity postulated in Proposition 2.1.

The integrals in (3.4) and (3.5) are Hellinger distances; cf. [3] for background and details. Were (3.4) to fail there would exist $\tilde\theta \ne \theta$ for which $g_\theta = g_{\tilde\theta}$ almost everywhere with respect to $\tilde Q$, which would violate identifiability in Assumption 1.1d); thus (3.4) holds. Finally, given $\tau > 0$ there is $\eta > 0$ such that $|h| < \eta$ implies that $\sup_{\theta,\mu}|g_{\theta+h}(\mu) - g_\theta(\mu)| < \tau$ (cf. the proof of Proposition 2.1). Consequently, since $x \to x^{1/2}$ is a uniformly continuous function, given $\varepsilon > 0$ there is $\delta > 0$ such that

$$\sup_{\theta,|h|\le\delta} |g_{\theta+h}(\mu)^{1/2} - g_\theta(\mu)^{1/2}| < \varepsilon^{1/2},$$

giving that the integral in (3.5) is at most ε, even uniformly in θ, establishing c).

Thus by [3, Theorem I.4.3] we have that $\hat\theta \to \theta$ almost surely under P_θ. By Assumption 1.1 c),

$$E_{\hat\theta}[M_{n+1}(dx)|N_{n+1}] = \frac{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1}) [dv/dv^*](x)}{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1})} v^*(dx),$$

and similarly for $E_\theta[M_{n+1}|N_{n+1}]$, so that

$$\|E_{\hat\theta}[M_{n+1}|N_{n+1}] - E_\theta[M_{n+1}|N_{n+1}]\| \tag{3.6}$$

$$= \int v^*(dx) \left| \frac{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1}) [dv/dv^*](x)}{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1})} - \frac{\int Q(dv) f_\theta(v) K(v,N_{n+1}) [dv/dv^*](x)}{\int Q(dv) f_\theta(v) K(v,N_{n+1})} \right|$$

$$\le \int v^*(dx) \left\{ \frac{\int Q(dv) |f_{\hat\theta}(v) - f_\theta(v)| K(v,N_{n+1}) [dv/dv^*](x)}{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1})} + \frac{\int Q(dv) f_\theta(v) K(v,N_{n+1}) [dv/dv^*](x)}{\int Q(dv) f_\theta(v) K(v,N_{n+1})} \frac{\int Q(dv) |f_{\hat\theta}(v) - f_\theta(v)| K(v,N_{n+1})}{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1})} \right\}$$

$$\le \sup_v |f_{\hat\theta}(v) - f_\theta(v)| \left\{ \frac{\int Q(dv) K(v,N_{n+1}) v(E)}{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1})} + \frac{\int Q(dv) f_\theta(v) K(v,N_{n+1}) v(E)}{\int Q(dv) f_\theta(v) K(v,N_{n+1})} \frac{\int Q(dv) K(v,N_{n+1})}{\int Q(dv) f_{\hat\theta}(v) K(v,N_{n+1})} \right\}$$

$$= (\sup_v |f_{\hat\theta}(v) - f_\theta(v)|) g_{\hat\theta}(N_{n+1})^{-1} \Big\{ \int Q(dv) K(v,N_{n+1}) v(E) + [\int Q(dv) K(v,N_{n+1})] E_\theta[M_{n+1}(E)|N_{n+1}] \Big\} .$$

Let us consider the components of the last expression in (3.6). By strong consistency and the equicontinuity assumption in Proposition 2.1, $\sup_v |f_{\hat\theta}(v) - f_\theta(v)| \to 0$ almost surely. The random variables $g_{\hat\theta}(N_{n+1})$ are consequently bounded away from zero in P_θ-probability (g_θ is the P_θ-density of the N_i); of course these random variables converge in no sense except in distribution. Next, we can write

$$\int \tilde Q(d\mu) \int Q(dv) K(v,\mu) v(E) = E[\tilde M(E)],$$

where $\tilde M$ is a random measure with law Q; this expectation is assumed finite, so therefore the sequence of i.i.d. random variables $\int Q(dv) K(v,N_{n+1}) v(E)$ is likewise bounded in probability. An essentially identical argument demonstrates the same of the sequence $(\int Q(dv) K(v,N_{n+1}))$. Finally, the assumption that $E_\theta[M(E)] < \infty$ implies that also the sequence $(E_\theta[M_{n+1}(E)|N_{n+1}])$ is bounded in P_θ-probability.

Combined, these observations confirm that (3.3) obtains. □

We examine strong consistency for pseudo-state estimators following a central limit theorem.

THEOREM 3.2. Assume that

a) for each v the function $\theta \to f_\theta(v)$ is twice continuously differentiable on Θ and that for some $\delta > 0$ the second derivative satisfies the uniform Lipschitz condition

$$\sup_v |f''_\theta(v) - f''_{\tilde\theta}(v)| \le c|\theta - \tilde\theta|^\delta \tag{3.7}$$

for all θ, $\tilde\theta$, where c is a positive constant;

b) there is $\delta > 0$ such that

$$\sup_\theta E_\theta[|f'_\theta(M)/f_\theta(M)|^{2+\delta}] < \infty \tag{3.8}$$

and

$$\sup_\theta E_\theta[|f''_\theta(M)/f_\theta(M)|^{1+\delta}] < \infty ; \tag{3.9}$$

c) for each θ,

$$E_\theta[M(E)^2] < \infty . \tag{3.10}$$

Then under P_θ,

$$n^{\frac{1}{2}}(E_{\hat\theta}[M_{n+1}|N_{n+1}] - E_\theta[M_{n+1}|N_{n+1}]) \xrightarrow{d} W , \tag{3.11}$$

where W is a random signed measure on E with the representation

$$W = Y\{E_\theta[(f'_\theta(M)/f_\theta(M))M|N] - E_\theta[f'_\theta(M)/f_\theta(M)|N]E_\theta[M|N]\} , \tag{3.12}$$

with Y independent of (N,M) and normally distributed with mean zero and variance

$$\sigma^2(\theta) = E_\theta[(g'_\theta(N)/g_\theta(N))^2]^{-1} . \tag{3.13}$$

PROOF. We break the argument into steps.

1) First, we establish that under P_θ,

$$n^{\frac{1}{2}}[\hat\theta - \theta] \xrightarrow{d} N(0,\sigma^2(\theta)); \tag{3.14}$$

our strategy -- similarly to the proof of Theorem 3.1 -- is to verify the following conditions of [3, Theorem I.8.1], where $\ell_\theta(\mu) = \log g_\theta(\mu)$.

i) $\theta \to \ell_\theta(\mu)$ is twice continuously differentiable for each μ;

ii) there is $\delta > 0$ such that

$$\sup_\theta E_\theta[|\ell'_\theta(N)|^{2+\delta}] < \infty ; \tag{3.15}$$

iii) there is $\eta > 0$ such that

$$\sup_\theta E_\theta[|\ell''_\theta(N)|^{1+\eta}] < \infty ; \tag{3.16}$$

iv) there is $\delta > 0$ such that for each θ and each compact subset Γ of Θ

$$E_\theta[\sup(|\ell''_\theta(N) - \ell''_{\bar\theta}(N)|/|\theta - \bar\theta|^\delta ; \bar\theta \in \Gamma)] < \infty .$$

We omit the straightforward verification that the statistical model (P_θ) is <u>regular</u> in the sense of [3, Section I.7]. Moreover, i) and iv) are consequences of (3.7), the former by an argument analogous to that used to prove Proposition 2.1 and the latter by direct computation.

To confirm (3.15) we observe that

$$E_\theta[|\ell'_\theta(N)|^{2+\delta}] = E_\theta[|E_\theta[f'_\theta(M)/f_\theta(M)|N]|^{2+\delta}] \le E_\theta[|f'_\theta(M)/f_\theta(M)|^{2+\delta}],$$

which is finite by (3.8). In the same way,

$$\begin{aligned} &E_\theta[|\ell''_\theta(N)|^{1+\eta}] \\ &= E_\theta[|E_\theta[f''_\theta(M)/f_\theta(M)|N] - E_\theta[f'_\theta(M)/f_\theta(M)|N]^2|^{1+\eta}] \\ &\le (E_\theta[|E_\theta[f''_\theta(M)/f_\theta(M)|N]|^{1+\eta}]^{1/(1+\eta)} \\ &\quad + E_\theta[|E_\theta[f'_\theta(M)/f_\theta(M)|N]^{2+2\eta}]^{1/(1+\eta)})^{1+\eta} ; \end{aligned}$$

this latter quantity is finite, provided that (3.7) and (3.8) hold, for $\eta = \delta/2$, and therefore (3.16) is fulfilled.

By [3, Theorem I.8.4] there is $\varepsilon > 0$ such that almost surely

$$n^{\frac{1}{2}}(1/\sigma^2(\theta))[\hat{\theta} - \theta] = n^{-\frac{1}{2}} \sum_{i=1}^{n} f_\theta'(N_i) + O(n^{-\varepsilon}) \tag{3.17}$$

uniformly on compact subsets of Θ. In particular, (3.17) implies (3.14). It also implies the corresponding law of the iterated logarithm, as we discuss further in Theorem 3.3.

2) As in (3.6), but now with more precision, we have

$$E_{\hat{\theta}}[M_{n+1}|N_{n+1}] - E_\theta[M_{n+1}|N_{n+1}] \tag{3.18}$$

$$= \frac{\int Q(dv)[f_{\hat{\theta}}(v) - f_\theta(v)]K(v,N_{n+1})v}{\int Q(dv) f_{\hat{\theta}}(v) K(v,N_{n+1})}$$

$$+ E_\theta[M_{n+1}|N_{n+1}] \frac{\int Q(dv)[f_\theta(v) - f_{\hat{\theta}}(v)]K(v,N_{n+1})}{\int Q(dv) f_\theta(v) K(v,N_{n+1})}$$

$$\doteq [\hat{\theta} - \theta]\left\{\frac{\int Q(dv) f_\theta'(v) K(v,N_{n+1})v}{\int Q(dv) f_\theta(v) K(v,N_{n+1})} - E_\theta[M_{n+1}|N_{n+1}] \frac{\int Q(dv) f_\theta'(v) K(v,N_{n+1})}{\int Q(dv) f_\theta(v) K(v,N_{n+1})}\right\}$$

$$+ o_p(|\hat{\theta} - \theta|^2)$$

[where $o_p(|\hat{\theta} - \theta|^2)/|\hat{\theta} - \theta|^2$ converges to zero in P_θ-probability as $n \to \infty$; this holds by virtue of (3.7)]

$$= [\hat{\theta} - \theta]\{E_\theta[(f_\theta'(M_{n+1})/f_\theta(M_{n+1}))M_{n+1}|N_{n+1}]$$

$$- E_\theta[f_\theta'(M_{n+1})/f_\theta(M_{n+1})|N_{n+1}]E_\theta[M_{n+1}|N_{n+1}]\},$$

where we have again applied (3.1).

3) For each n the two factors in the last expression in (3.18) are independent and the distribution of the latter is independent of n. Together with (3.14), these observations imply that (3.11) holds, with W having the structure given by (3.12). Note that in view of (3.8) and (3.10) W is a legitimate signed measure on E almost surely and that the convergence in distribution in (3.11) holds with respect to the variation norm topology. □

To conclude the section we return to consistency of pseudo-state estimators, but now in the almost sure sense. Of several possible supplementary hypotheses, we choose what seems the simplest.

THEOREM 3.3. Suppose that the hypotheses of Theorem 3.2 are fulfilled and that for each θ there is a constant c, allowed to depend on θ, such that

$$P_\theta\{N(E) \le c\} = 1. \tag{3.19}$$

Then almost surely under P_θ,

$$\| E_{\hat{\theta}}[M_{n+1}|N_{n+1}] - E_\theta[M_{n+1}|N_{n+1}] \| \to 0 .$$

PROOF. By (3.17), (3.18) and (3.19), within error converging to zero almost surely

$$\| E_{\hat\theta}[M_{n+1} | N_{n+1}] - E_\theta[M_{n+1} | N_{n+1}] \| \tag{3.20}$$

$$\le |\hat\theta - \theta| \{ \| E_\theta[(f_\theta'(M_{n+1})/f_\theta(M_{n+1}))M_{n+1} | N_{n+1}] \|$$

$$+ |E_\theta[f_\theta'(M_{n+1})/f_\theta(M_{n+1}) | N_{n+1}]| \cdot \| E_\theta[M_{n+1} | N_{n+1}] \| \}$$

$$\le 2c|\hat\theta - \theta| E_\theta[|f_\theta'(M_{n+1})/f_\theta(M_{n+1})| \,|N_{n+1}].$$

That the set functions appearing here are signed measures almost surely was observed in the course of proving Theorem 3.2. As remarked during that proof, (3.17) implies that the maximum likelihood estimators $\hat\theta$ satisfy a law of the iterated logarithm: as $n \to \infty$

$$|\hat\theta - \theta| = O(\left[\frac{\log\log n}{n}\right]^{\frac{1}{2}}) \tag{3.21}$$

almost surely. From (3.8), with δ the exponent appearing there,

$$P_\theta\{E_\theta[|f_\theta'(M_{n+1})/f_\theta(M_{n+1})| \,|N_{n+1}] > (n/\log n)^{\frac{1}{2}}\}$$

$$\le [(\log n)/n]^{1+\delta/2} E_\theta[E_\theta[|f_\theta'(M_{n+1})/f_\theta(M_{n+1})| \,|N_{n+1}]^{2+\delta}]$$

$$\le [(\log n)/n]^{1+\delta/2} E_\theta[|f_\theta'(M)/f_\theta(M)|^{2+\delta}].$$

Consequently, by the usual Borel-Cantelli argument,

$$E_\theta[|f_\theta'(M_{n+1})/f_\theta(M_{n+1})| \,|N_{n+1}] \le [n/(\log n)]^{\frac{1}{2}}, \tag{3.22}$$

for all sufficiently large n, almost surely.

Together (3.20) - (3.22) yield

$$\| E_{\hat\theta}[M_{n+1} | N_{n+1}] - E_\theta[M_{n+1} | N_{n+1}] \| = O(\left[\frac{\log\log n}{n}\right]^{\frac{1}{2}}) \tag{3.23}$$

almost surely, which completes the proof. $\square$

With stronger moment assumptions on $f_\theta'(M)/f_\theta(M)$ one can establish rates of convergence sharper than (3.23).

4. Examples

In this section we illustrate our procedures, mainly the computational techniques developed in Section 2, for two specific cases.

EXAMPLE 4.1. Random exponential densities. Suppose that $E = \mathbb{R}_+$ and that the directing measure M has the form

$$M(dx) = Ye^{-Yx}\,dx, \tag{4.1}$$

where Y is a nonnegative random variable. That is, M "is" an exponential density with random parameter Y. As statistical model we take $\{P_\theta : \theta > 0\}$, where under P_θ, Y itself is exponentially distributed with parameter θ. For notational ease let $R_i = N_i(\mathbb{R}_+)$ and $A_i = \int_0^\infty x\, N_i(dx)$.

Assumptions 1.1 are satisfied, and the log-likelihood function corresponding to observation of i.i.d. Cox processes $N_1,\dots,N_n$ (with directing measures $M_i(dx) = Y_i e^{-Y_i(x)} dx$) is

$$L_n(\theta) = \sum_{i=1}^n \log(\int_0^\infty \theta e^{-\theta t} t^{R_i} e^{-tA_i}\,dt)$$

$$= \sum_{i=1}^n \log(\theta(R_i!)/(\theta + A_i)^{R_i+1}).$$

The resultant likelihood equation, a special case of (2.1), is

$$\theta = n/\sum_{i=1}^{n} ((R_i+1)/(\theta+A_i)) = n/\sum_{i=1}^{n} E_\theta[Y_i|N_i], \tag{4.2}$$

where we have used (1.3). Despite its being analogous to the likelihood equation $\theta = n/\sum_{i=1}^{n} Y_i$ associated with observation (not permitted in our setting) of $Y_1,\ldots,Y_n$, (4.2) is, rather apparently, difficult to solve explicitly.

Fortunately the EM algorithm works very effectively in this situation. With h_θ the exponential density function with parameter θ and

$$\lambda_n(\theta) = \sum_{i=1}^{n} \log h_\theta(Y_i) = n \log \theta - \theta \sum_{i=1}^{n} Y_i,$$

the unobservable log-likelihood for $Y_1,\ldots,Y_n$, we have as $P_{\tilde\theta}$-pseudo-log-likelihood function

$$\begin{aligned} E_{\tilde\theta}[\lambda_n(\theta)|G_n] &= \sum_{i=1}^{n} E_{\tilde\theta}[\log h_\theta(Y_i)|N_i] \\ &= \sum_{i=1}^{n} E_{\tilde\theta}[\log \theta - \theta Y_i|N_i] \\ &= n \log \theta - \theta \sum_{i=1}^{n} E_{\tilde\theta}[Y_i|N_i]. \end{aligned}$$

As a function of θ, this is evidently maximized at the solution θ to the likelihood equation corresponding to (2.6), i.e.,

$$\theta = n/\sum_{i=1}^{n} E_{\tilde\theta}[Y_i|N_i] = n/\sum_{i=1}^{n} ((R_i+1)/(\tilde\theta+A_i)). \tag{4.3}$$

Comparison of (4.2) and (4.3) confirms that the latter is the obvious

iterative method for solving the former. Moreover, (4.2) has a unique solution $\hat\theta$ maximizing $L_n(\cdot)$, and Theorem 2.4 implies that the sequence of iterates produced by the EM algorithm converges to $\hat\theta$ regardless of the initializing value.

To illustrate the procedure explicitly we present in Table 1 a numerical example. For various values of θ and the sample size n we give the estimator $\hat\theta$ computed via the EM algorithm, the number of iterations of (4.3) required for convergence (within 0.01; the algorithm was always initialized with $\theta^0 = 1$) and the maximum likelihood estimator $\hat\theta^* = n/\sum_{i=1}^{n} Y_i$ corresponding to the unobservable random variables $Y_1,\ldots,Y_n$.

<u>Table 1 near here</u>

We also compare, but only for the case $\theta = 5$, $n = 50$, various state estimators for M_{51}. From (1.3), the true state estimators have the form

$$E_\theta[M|N](dx) = [(R+1)(\theta+A)^{R+1}/(\theta+A+x)^{R+2}]dx, \tag{4.4}$$

where $R = N(\mathbb{R}_+)$ and $A = \int_0^\infty yN(dy)$. The entries in Table 2 are values of $M_{51}(dx)/dx = Y_{51}e^{-Y_{51}x}$, the true state estimator $E_{\theta=5}[Y_{51}e^{-Y_{51}x}|N_{51}]$ and the pseudo-state estimator $E_{\hat\theta=4.79}[Y_{51}e^{-Y_{51}x}|N_{51}]$ for a number of x-values.

<u>Table 2 near here</u>

Our second example is treated nonparametrically in [6].

EXAMPLE 4.2. <u>Mixed Poisson processes</u>. Suppose that under P_θ the directing measures are

$$N_i(dx) = Y_i h_\theta(x)\nu(dx), \tag{4.4}$$

with ν a diffuse probability measure on E, where for each θ, h_θ is a nonnegative function on E satisfying $\int h_\theta d\nu = 1$ (this is required in order to ensure identifiability), and where Y is a positive random variable with P_θ-density function f_θ. Although this model lies slightly outside the structure set forth in Section 1 — the domination hypotheses fail — we can apply the same kinds of methods. However the unobservable log-likelihood function is now that for $(N_1,Y_1),\ldots,(N_n,Y_n)$ rather than $N_1,\ldots,N_n$.

The observable n-sample log-likelihood function is

$$L_n(\theta) = \sum_{i=1}^{n} \log\{\int_0^\infty f_\theta(t)\exp[\int_E(1-th_\theta)d\nu + \int_E(\log t + \log h_\theta)dN_i]dt\}$$

$$= n + \sum_{i=1}^{n} \log\{\int_0^\infty f_\theta(t)e^{-t}t^{R_i}dt\} + \sum_{i=1}^{n}\int_E(\log h_\theta)dN_i,$$

where $R_i = N_i(E)$. Under appropriate differentiability hypotheses on f_θ and h_θ the associated likelihood equation is

$$-\sum_{i=1}^{n}\int_E(h_\theta'/h_\theta)dN_i = \sum_{i=1}^{n} E_\theta[f_\theta'(Y_i)/f_\theta(Y_i)|N_i] \tag{4.5}$$

$$= \sum_{i=1}^{n} \frac{\int f_\theta'(t)e^{-t}t^{R_i}dt}{\int f_\theta(t)e^{-t}t^{R_i}dt}.$$

No matter how simple the f_θ and h_θ, (4.5) is formidable if not impossible to solve; thus we are led again to the EM algorithm, although now it may be less helpful than in Example 4.1.

With $\Lambda_n(\theta)$ the log-likelihood function for $(N_1,Y_1),\ldots,(N_n,Y_n)$ (we omit

an explicit computation) the pseudo-log-likelihood function under $P_{\tilde\theta}$ is

$$E_{\tilde\theta}[\Lambda_n(\theta)|G_n] \tag{4.6}$$

$$= \sum_{i=1}^{n} E_{\tilde\theta}[\log f_\theta(Y_i)|N_i] + n + \sum_{i=1}^{n} E_{\tilde\theta}[Y_i|N_i]$$

$$+ \sum_{i=1}^{n} R_i E_{\tilde\theta}[\log Y_i|N_i] + \sum_{i=1}^{n}\int_E(\log h_\theta)dN_i.$$

In (4.6) only the first and fifth terms on the right-hand side depend on θ; thus the likelihood equation is

$$0 = (d/d\theta)E_{\tilde\theta}[\Lambda_n(\theta)|G_n] \tag{4.7}$$

$$= \sum_{i=1}^{n}\{E_{\tilde\theta}[f_\theta'(Y_i)/f_\theta(Y_i)|N_i] + \int_E(h_\theta'/h_\theta)dN_i\}$$

$$= \sum_{i=1}^{n}\{\frac{\int f_{\tilde\theta}(t)e^{-t}t^{R_i}[f_\theta'(t)/f_\theta(t)]dt}{\int f_{\tilde\theta}(t)e^{-t}t^{R_i}dt} + \int_E(h_\theta'/h_\theta)dN_i\}.$$

It is not obvious that (4.7) is materially more tractable than (4.5). Of course given sufficient special structure for the f_θ and h_θ, (4.7) can be solved explicitly for θ (although so possibly might (4.5)); however we do not pursue such cases here.

Table 1

Parameter Estimates for Example 4.1

θ	n	$\hat{\theta}$ via EM	Iterations	$\hat{\theta}^* = 1/\bar{Y}$
2	10	1.24	7	1.47
	50	1.89	14	1.88
	100	1.59	10	2.09
	1000	1.98	12	1.97
5	10	2.82	12	6.87
	50	4.57	12	4.90
	100	4.79	16	5.28
	1000	4.92	18	5.12
10	10	11.89	33	12.91
	50	8.98	19	11.59
	100	10.68	20	9.19
	1000	8.24	20	9.68

Table 2

State Estimators for Example 4.1

x	$X_{51}(x) = Y_{51}e^{-Y_{51}x}$	$E_{\theta=5}[X_{51}(x) \mid N_{51}]$	$E_{\hat{\theta}=4.79}[X_{51}(x) \mid N_{51}]$
0	.235	.234	.238
.25	.222	.217	.220
.50	.209	.201	.204
.75	.197	.186	.189
1.00	.186	.173	.175
1.50	.165	.150	.152
2.00	.147	.131	.132
2.50	.131	.115	.115
3.00	.116	.101	.101
4.00	.092	.080	.079
5.00	.073	.063	.063
7.50	.040	.037	.037
10.00	.022	.023	.023

References

[1] P.J. Bickel and K.A. Doksum, Mathematical Statistics (Holden-Day, San Francisco, 1977).

[2] A.P. Dempster, N.M. Laird and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm. J. Roy. Statist. Soc. B 39 (1977) 1-38.

[3] I.A. Ibragimov and R.Z. Has'minskii, Statistical Estimation: Asymptotic Theory (Springer-Verlag, New York, 1981).

[4] O. Kallenberg, Random Measures, 3rd edition (Akademie-Verlag, Berlin, and Academic Press, New York, 1983).

[5] A.F. Karr, State estimation for Cox processes on general spaces. Stochastic Process. Appl. 14 (1983) 209-232.

[6] A.F. Karr, Combined nonparametric inference and state estimation for mixed Poisson processes. Z. Wahrsch. und verw. Geb. 66 (1984) 81-96.

[7] A.F. Karr, State estimation for Cox processes with unknown probability law. Stochastic Process. Appl. 19 (1985) 115-131.

[8] A.F. Karr, Point Processes and their Statistical Inference (Marcel Dekker, New York, in press).

[9] B.G. Lindsay, The geometry of mixture likelihoods, I. A general theory and II. The exponential family. Ann. Statist. 11 (1983) 86-94 and 783-792.

[10] D.L. Snyder, Random Point Processes. (Wiley-Interscience, New York, 1975).

[11] Y. Vardi, L.A. Shepp and L. Kaufman, A statistical model for positron emission tomography. J. Amer. Statist. Assoc. 80 (1985) 8-20.

[12] C.F.J. Wu, On the convergence properties of the EM algorithm. Ann. Statist. 11 (1983) 95-103.

END
FILMED

5-86

DTIC